

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 9, 2005, 21:24:10 ; Search time 3569 Seconds  
(without alignments)  
12575.926 Million cell updates/sec

Title: US-10-717-580-11

Perfect score: 7582

Sequence: 1 caacaatcggatataaac.....tattaaatttcgactgtg 7582

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: Genesecq16Dec04:\*  
2: Genesecq1980s:\*  
3: Genesecq1990s:\*  
4: Genesecq2000s:\*  
5: Genesecq2001as:\*  
6: Genesecq2001bs:\*  
7: Genesecq2002as:\*  
8: Genesecq2002bs:\*  
9: Genesecq2003as:\*  
10: Genesecq2003bs:\*  
11: Genesecq2003cs:\*  
12: Genesecq2004as:\*  
13: Genesecq2004bs:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7489.6	98.8	7582	3	AAS59215
2	7487.4	98.8	7582	2	AAS59215 Human end
3	6448	85.0	56093	6	AAX25665 Complete
4	6436	84.9	10499	3	ABL61744 Colon ade
5	6124.6	80.8	8523	7	ABN97929 Human ret
6	5108	67.4	9502	10	ADS30988 Human gen
7	3738.4	49.3	6394	5	Adf59718 Human con
8	3431	45.3	4349	5	AAS84210 DNA encod
9	3309.6	43.7	7466	5	AAS76475 DNA encod
10	3049.8	40.2	22436	9	AAS68626 DNA encod
11	3049.8	40.2	22436	10	ADA02882 Human PAP
12	3049.8	40.2	22436	10	Adb72620 Human PAP
13	2948	38.9	139573	12	Adc85361 Mouse pap
14	2948	38.9	139573	10	Adm74477 Human car
15	2948	38.9	156416	13	Adh58564 Human Na+
16	2932	38.7	5154	5	ABD32817 Human can
17	2908.2	38.4	3464	10	Aas67609 DNA encod
18	2889.6	38.1	2930	6	Ade09587 Novel DNA
19	2888.6	38.1	2946	6	Aad34195 Human syn
20	2888.6	38.1	2946	3	Aax77526 Human sec
					Aaz59468 Human sec

21	2888.6	38.1	2946	10	ADC38776
22	2734.4	36.1	3831	5	AAS71727
23	2688.6	35.5	2938	2	AAX25656 Human end
24	2688.6	35.5	2938	3	AAS59206 Gag and p
25	2670.8	35.2	173564	13	ABD32953 Human can
26	2613.2	34.5	3372	2	AAX25663 Human end
27	2613.2	34.5	3372	3	AAX59213 Partial p
28	2550.4	33.6	2599	3	AAS76205 DNA encod
29	2541.6	33.5	4535	5	Adc32196 Human nov
30	2541.6	33.5	4535	10	ADC32196
31	2368.4	31.2	8279	5	AAS76474 DNA encod
32	2282.4	30.1	2784	3	ABN97930 Human ret
33	2251.8	29.7	2372	2	AAX25664 Human end
34	2251.8	29.7	2372	3	AAS59214 R-U5 regi
35	2241	29.6	8294	5	AAS84209 DNA encod
36	2241	29.6	8294	10	ADP59732 Human con
37	2220.4	29.3	161334	11	ACN44334 Human gen
38	2216	29.2	46340	3	ABN97978 Human ret
39	2204.6	29.1	2782	2	AAX25661 Human end
40	2204.6	29.1	2782	3	AAS59211 5' non co
41	2204.6	29.1	2782	5	AAX20069 HERV-W en
42	2201.4	29.0	2781	5	AAF55630 Nucleotid
43	2182.8	28.8	3903	5	AHH46211 Phosphori
44	2168	28.6	8399	7	ADS30993 Human gen
45	2131.4	28.1	2575	2	AAX25668 Human end

## ALIGNMENTS

### RESULT 1

AAS59215

ID AAS59215 standard; DNA; 7582 BP.

AC AAS59215;

DT 07-NOV-2000 (first entry)

XX Human endogenous retrovirus W (HERV-W) sequence.

XX Autoimmune disease; retrovirus; human endogenous retrovirus W; HERV-W;  
gag gene; pregnancy; multiple sclerosis; T cell proliferation; ss.

OS Human endogenous retrovirus.

XX	Key	Location/Qualifiers
FT	LTR	1..120
FT		/*tag= a
FT		/note= "R of 5' LTR"
FT		121..575
FT		/*tag= b
FT	primer_bind	/note= "U5 of 5' LTR"
FT		579..596
FT		/*tag= c
FT	CDS	5581..7194
FT		/*tag= d
FT		/note= "ORF1 env538"
FT	CDS	7039..7194
FT		/*tag= e
FT		/note= "ORE2 52 AA"
FT	CDS	7112..7255
FT		/*tag= f
FT		/note= "ORF3 48 AA"
FT	misc_feature	7244..7254
FT		/*tag= g
FT		/note= "polypurine tract"
FT	LTR	7256..7582
FT		/*tag= h
FT		/note= "U3-R of 3' LTR"
FT	polya_signal	7563..7569
FT		/*tag= i
FN		WO200043521-A2.

GenCore version 5.1.6  
Copyright (C) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 15, 2005, 23:53:58 ; Search time 31202 Seconds  
(without alignments)  
11774.484 Million cell updates/sec

Title: US-10-717-580-11  
Perfect score: 7582  
Sequence: 1 caacaatcgggataataaacc.....tattaatcttgcacrtgcr 7582

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0  
Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*  
1: gb\_ba.\*  
2: gb\_hgt.\*  
3: gb\_in.\*  
4: gb\_om.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vt.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7489.6	98.8	7582	6	BD196245 Endogenet
2	7489.6	98.8	7582	6	BD267487 Endogenet
3	7489.6	98.8	7582	6	AX000966 Sequence
4	7489.6	98.8	7582	6	AX027480 Sequence
5	6448	85.0	10222	9	AY101582 Homo sapi
6	6448	85.0	10222	9	AY101585 Homo sapi
7	6448	85.0	56093	6	AX329572 Sequence
8	6448	85.0	56093	9	HSAC000064
9	6448	85.0	149194	9	AC007566 Homo sapi
10	6446.4	85.0	10222	9	AY101583 Homo sapi
11	6440	84.9	10222	9	AY101584 Homo sapi
12	6436	84.9	10499	6	BD221808 Nucleic s
13	6436	84.9	10499	6	AX007980 Sequence
14	6346.4	83.7	10229	9	AY101586 Pan trogl
15	6344.8	83.7	10229	9	AY101587 Pan trogl
16	6344.8	83.7	184575	9	AC145964 Pan trogl
17	6269.6	82.7	10230	9	AY101588 Gorilla g
18	6269.6	82.7	10230	9	AY101589 Gorilla g
19	6037.8	79.6	10122	9	AY101590 Pongo pyg

20	6025.8	79.5	10124	9	AY101591 Pongo pyg
21	5956.8	78.6	10246	9	AY101593 Hylobates
22	5950.4	78.5	10248	9	AY101592 Hylobates
23	5824.2	76.8	158033	9	AC018926 Homo sapi
24	5492	72.4	105989	9	AC008121 Homo sapi
25	5492	72.4	110000	2	AC009727_2 Continuation (3 of
26	4755.6	62.7	117321	2	AC092510 Papio anu
27	4302.2	56.7	114621	9	AC005187 Homo sapi
28	4295.8	56.7	46575	2	AC080036 Homo sapi
29	4162	54.9	111140	2	AL592310 Human DNA
30	3944.6	52.0	183680	9	AC098859 Homo sapi
31	3779.8	49.9	73070	9	AL590143 Human DNA
32	3731.6	49.2	173788	9	AC022555 Homo sapi
33	3731.6	49.2	186540	9	AC090341 Homo sapi
34	3731.6	49.2	187837	9	AC079065 Homo sapi
35	3716.6	49.0	160883	2	AL158814 Homo sapi
36	3537.2	46.7	137947	9	HS4514 Human DNA
37	3442.8	45.4	153444	2	AC040948 Homo sapi
38	3442.8	45.4	205035	9	CNS000008 Human chr
39	3438.2	45.3	101846	9	HS19711 Human DNA
40	3438.2	45.3	224629	2	AC139260 Homo sapi
41	3420.6	45.1	68693	9	AC108746 Homo sapi
42	3367.4	44.4	118312	9	AC055713 Homo sapi
43	3367.4	44.4	197549	2	AC145423 Homo sapi
44	3300.6	43.5	169250	9	AP002812 Homo sapi
45	3290.2	43.4	134412	2	AP002377 Homo sapi

## ALIGNMENTS

RESULT 1  
BD196245  
LOCUS  
DEFINITION  
BD196245  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

BD196245 7582 bp DNA linear PAT 17-JUL-2003  
Endogenetic retroviral sequences, associated with autoimmune diseases and/or with pregnancy disorders.  
BD196245.1 GI:33006015  
JP 2002512530-A/11.  
unidentified  
unclassified.  
1 (bases 1 to 7582)  
Beseme,F., Blond,J.L., Bouton,O., Mandrand,B. and Mallet,F.  
Endogenetic retroviral sequences, associated with autoimmune diseases and/or with pregnancy disorders  
Patent: JP 2002512530-A 11 23-APR-2002;  
BIO MERIEUX  
OS Unidentified  
PN JP 2002512530-A/11  
PD 23-APR-2002  
PF 06-JUL-1998 JP 1999508244  
PR 07-JUL-1997 FR 97/08815  
PI FREDERIC BESEME,JEAN LUC BLOND,OLIVIER BOUTON,BERNARD MANDRAND,  
PI FRANCOIS MALLET  
PC C12N15/48,C07K14/15,C12Q16/68,C07K16/10,G01N33/569 CC  
Strandedness: Single;  
CC Topology: Linear;  
CC Endogenetic retroviral sequences, associated with autoimmune diseases  
CC and/or with pregnancy disorders  
FH key Location/Qualifiers  
FT source 1..7582 /organism='Unidentified'.  
location/Qualifiers  
1..7582 /organism='unidentified'  
/mol\_type='genomic DNA'  
/db\_xref='taxon:32644'

Query Match 98.8%; Score 7489.6; DB 6; Length 7582;

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 15, 2005, 22:59:08 ; Search time 21789 Seconds  
(without alignments)  
13245.366 Million cell updates/sec

Title: US-10-717-580-11

Perfect score: 7582

Sequence: 1 caacaatcggtataaacc.....tattaaatcttgcaatgcgc 7582

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: gb\_est1.\*

2: gb\_est2.\*

3: gb\_hcc.\*

4: gb\_est3.\*

5: gb\_est4.\*

6: gb\_est5.\*

7: gb\_est6.\*

8: gb\_gss1.\*

9: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2180.4	28.8	2748	3	CR605851 full-leng
2	2176.4	28.7	2749	3	CR617248 full-leng
3	2174.4	28.7	2500	3	CR622175 full-leng
4	2169.4	28.6	2748	3	CR625046 full-leng
5	2147.4	28.3	2716	3	CR613169 full-leng
6	969	12.8	1071	5	CR6365066 full-leng
7	921.2	12.1	998	5	CR6337769 full-leng
8	909.6	12.0	1058	5	CR6378303 full-leng
9	908.6	12.0	935	5	CR6391741 full-leng
10	885.2	11.7	955	5	CR6347952 full-leng
11	885	11.7	1019	5	CR6349636 full-leng
12	866.6	11.4	1014	5	CR6378304 full-leng
13	848.8	11.2	6400	3	CR649521 full-leng
14	834.8	11.0	924	5	CR6347314 full-leng
15	820.8	10.8	931	5	CR6389656 full-leng
16	820	10.8	903	5	CR6408752 full-leng
17	812	10.7	872	5	CR6326647 full-leng
18	804	10.6	883	5	CR6408733 full-leng
19	803.8	10.6	966	5	CR6380176 full-leng
20	789.2	10.4	877	5	CR6347619 full-leng
21	776.2	10.2	921	5	CR6408734 full-leng
22	771	10.2	836	5	CR6388766 full-leng
23	768.6	10.1	995	5	CR6389657 full-leng
24	765.4	10.1	844	5	CR6347111 full-leng

25 764.8 10.1 828 5 BX367907  
c 26 753.8 9.9 846 5 BX368078  
c 27 751.6 9.9 900 5 BX459153  
c 28 724.6 9.6 758 5 BX357208  
29 723.4 9.5 879 5 BX325902  
30 719.4 9.5 924 5 BX409328  
31 709.4 9.4 771 4 BX087886  
32 697.2 9.2 723 1 AU138405  
33 677.4 8.9 702 1 AU138097  
34 673.6 8.9 795 8 BZ611989  
35 665.6 8.8 714 2 BE734284  
36 665 8.8 719 7 CN272394  
37 649.8 8.6 944 5 BX368759  
c 38 639.6 8.4 744 8 BZ601435  
c 39 637.8 8.4 833 8 BZ607553  
c 40 621.6 8.2 658 5 BX430050  
41 621.6 8.2 851 7 CR765248  
c 42 617 8.1 776 8 AQ351025  
43 616.2 8.1 738 9 HSA408776  
44 615.2 8.1 733 5 BX374329  
45 612 8.1 792 5 BX409304

## ALIGNMENTS

RESULT 1  
CR605851  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REMARK  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

2748 bp mRNA linear HTC 21-JUL-2004  
full-length cDNA clone CSODE012YJ24 of Placenta of Homo sapiens (human)  
CR605851.1 GI:50486658  
HTC; CNSLT CDNA.  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
Li, W.B., Gruber, C., Jesses, J. and Polayes, D.  
Full-length cDNA libraries and normalization  
Unpublished  
Contact : Feng Liang Email : fliang@lifetech.com URL :  
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
Faraday Avenue  
2 (bases 1 to 2748)  
Genoscope.  
Direct Submission  
Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)  
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime  
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library  
was normalized. Library was constructed by Life Technologies, a  
division of Invitrogen.  
Location/Qualifiers  
1. 2748  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CSODE012YJ24"  
/tissue type="Placenta"  
/plasmid="pCMVSPORT\_6"

COMMENT  
FEATURES  
source

28.8%; Score 2180.4; DB 3; Length 2748;  
Query Match  
Best Local Similarity 98.4%; Pred. No. 0;  
Matches 2171; Conservative 34; Mismatches 0; Indels 1; Gaps 1;

ORIGIN

5352 GATACATCTCGGAAGACCTACCCAGTCATTTATYATCCCACTGGGTTAAAGTG 5411  
|||||  
544 GATACATCTCGGAAGACCTACCCAGTCATTTATCTACCCCACTGGGTTAAAGTG 603  
|||||